

CLASSIFICATION OF PUBLIC SENTIMENT TOWARDS STUNTING PREVENTION PROGRAM USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE ON X APPLICATION

Desi Fitria¹, Ryan Randy Suryono*²

^{1,2}Information Systems, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

Email: ¹desi_fitria@teknokrat.ac.id, ²ryan@teknokrat.ac.id

(Article received: October 12, 2024; Revision: October 31, 2024; published: December 29, 2024)

Abstract

Indonesia has serious health problems, one of which is stunting among children. Stunting is caused by chronic malnutrition that affects a child's physical and cognitive growth. To address its impact, the government launched a prevention program that focuses on improving nutrition, improving sanitation, and health education. Public response to these programs has varied, with some supportive and others skeptical. In the digital age, public opinion is expressed through social media, making sentiment analysis important to understand public perception. This research aims to classify public sentiment towards the stunting prevention program using Naive Bayes and Support Vector Machine (SVM) methods. Data preprocessing includes cleaning, case folding, tokenizing, stopwords, and stemming, ensuring the text data is ready for analysis. The dataset consists of 5907 tweets divided by a ratio of 80:20, resulting in 4725 tweets for training data and 1182 tweets for testing data. The analysis results show that the Naive Bayes model achieved an accuracy of 95.34% for training data and 84.52% for testing data, while SVM achieved an accuracy of 95.43% for training data and 96.74% for testing data, indicating the performance of the SVM model is better than the Naive Bayes model. The important impact of this research is to assist policymakers in understanding the public's perception of government programs so that they can design communication strategies.

Keywords: Sentiment Analysis; Naive Bayes; Stunting Program; Support Vector Machine

KLASIFIKASI SENTIMEN PUBLIK TERHADAP PROGRAM PENCEGAHAN STUNTING MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE PADA APLIKASI X

Abstrak

Indonesia memiliki masalah kesehatan yang serius, salah satunya adalah stunting pada anak-anak. Stunting disebabkan oleh kekurangan gizi kronis yang mempengaruhi pertumbuhan fisik dan kognitif anak. Untuk mengatasi dampaknya, pemerintah meluncurkan program pencegahan yang berfokus pada peningkatan gizi, perbaikan sanitasi, dan edukasi kesehatan. Tanggapan masyarakat terhadap program ini bervariasi, dengan beberapa mendukung dan yang lain skeptis. Dalam era digital, opini publik diekspresikan melalui media sosial, sehingga analisis sentimen menjadi penting untuk memahami persepsi masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap program pencegahan stunting menggunakan metode *Naive Bayes* dan *Support Vector Machine* (SVM). Proses preprocessing data meliputi *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*, memastikan data teks siap untuk analisis. Dataset terdiri dari 5907 tweet yang dibagi dengan rasio 80:20, menghasilkan 4725 tweet untuk data training dan 1182 tweet untuk data testing. Hasil analisis menunjukkan bahwa model *Naive Bayes* mencapai akurasi 95,34% untuk data training dan 84,52% untuk data testing, sementara SVM mencapai akurasi 95,43% data training dan 96,74% untuk data testing, menunjukkan performa model SVM lebih baik dari model *Naive Bayes*. Dampak penting dari penelitian ini adalah membantu para pengambil kebijakan dalam memahami persepsi masyarakat terhadap program-program pemerintah sehingga dapat merancang strategi komunikasi.

Kata kunci: Analisis Sentimen; Naive Bayes; Program Stunting; dan Support Vector Machine

1. PENDAHULUAN

Di Indonesia, stunting merupakan salah satu masalah kesehatan yang mendesak, terutama di kalangan anak-anak [1]. Program pencegahan stunting telah diluncurkan oleh pemerintah dengan harapan dapat menekan prevalensi stunting [2]. Meskipun program ini bertujuan untuk meningkatkan kualitas kesehatan generasi mendatang, tanggapan masyarakat terhadap program ini beragam. Sebagian mendukung langkah-langkah yang diambil pemerintah, sementara yang lain merasa skeptis atau bahkan tidak puas dengan pelaksanaan program tersebut [3]. Dalam era digital saat ini, opini publik sering kali diekspresikan melalui media sosial, forum online, dan platform berita. Oleh karena itu, analisis sentimen otomatis menjadi penting untuk mengukur persepsi publik terhadap program ini [4]. Dengan mengklasifikasikan opini publik menjadi positif, negatif, atau netral, metode ini memungkinkan kita untuk memahami secara mendalam tingkat penerimaan dan efektivitas program di masyarakat [5].

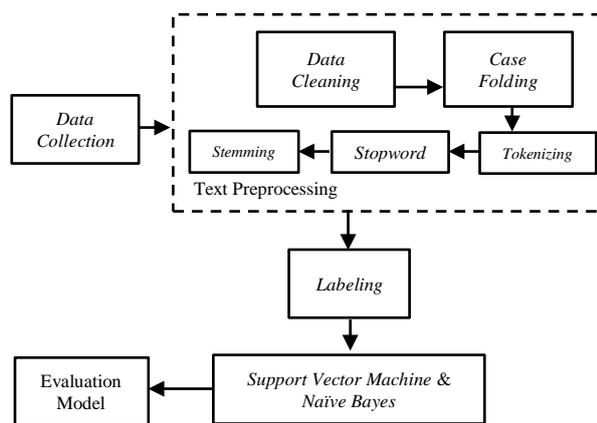
Salah satu pendekatan yang banyak digunakan untuk menganalisis sentimen adalah metode berbasis machine learning, seperti *Naive Bayes* dan *Support Vector Machine (SVM)* [6]. Penelitian terdahulu membahas tentang *Autism Spectrum Disorder (ASD)* dan menggunakan *machine learning* untuk mendeteksi ASD secara otomatis. Penelitian ini membandingkan dua algoritma, yaitu *Naive Bayes* dan *Support Vector Machine (SVM)*, dan menunjukkan bahwa *Naive Bayes* memiliki performa yang lebih unggul dengan akurasi 96,45%, dibandingkan SVM yang hanya mencapai 81,56%. Dengan hasil tersebut, *Naive Bayes* dianggap lebih efektif dalam mendeteksi ASD [7]. Analisis kasus stunting balita menggunakan algoritma SVM untuk mengklasifikasikan kasus. Dataset yang digunakan terdiri dari 6500 data yang diambil dari platform Kaggle. Setelah dilakukan proses *preprocessing* data, SVM dengan kernel linear menghasilkan akurasi sebesar 82%, dengan presisi 80%, dan recall 86%. Hal ini menunjukkan bahwa SVM juga merupakan algoritma yang dapat diandalkan untuk analisis klasifikasi dalam kasus kesehatan seperti stunting [8].

Penelitian tentang masalah stunting pada balita di Kecamatan Poto Tano, yang menunjukkan risiko tinggi pada populasi tersebut. Penelitian ini menggunakan dataset dari Puskesmas yang berjumlah 5829 data gizi anak untuk mendeteksi stunting. Dengan menggunakan metode *Naive Bayes*, penelitian ini berfokus pada prediksi risiko stunting berdasarkan faktor-faktor gizi anak. *Naive Bayes* digunakan karena kemampuannya dalam menangani data probabilistik yang besar dan kompleks, terutama dalam permasalahan yang berkaitan dengan kesehatan masyarakat [9]. Kebijakan pemindahan ibu kota Indonesia (IKN)

memicu respons di media sosial X. Penelitian ini menganalisis 5128 tweet tentang IKN, membandingkan algoritma SVM dan *Naive Bayes*. Dengan optimasi SMOTE, SVM mencapai akurasi 84% dan presisi 82%, sedangkan *Naive Bayes* mencapai akurasi 77% dan presisi 71%. Hasil menunjukkan SVM lebih unggul dalam analisis sentimen positif [10]. Penelitian-penelitian ini menunjukkan bahwa baik *Naive Bayes* maupun SVM memiliki keunggulan masing-masing dalam analisis klasifikasi untuk isu-isu kesehatan, seperti *Autism Spectrum Disorder (ASD)* dan stunting pada balita. Dengan mengeksplorasi lebih dalam faktor-faktor yang memengaruhi stunting dan menggunakan analisis sentimen yang komprehensif, penelitian ini mengatasi kekurangan dari penelitian sebelumnya dan memberikan wawasan yang berguna bagi pembuat kebijakan dalam merancang strategi intervensi yang lebih tepat sasaran [11]–[13].

Berdasarkan latar belakang masalah tersebut menunjukkan bahwa baik *Naive Bayes* maupun SVM memiliki keunggulan masing-masing dalam analisis klasifikasi untuk isu-isu kesehatan, seperti *Autism Spectrum Disorder (ASD)* dan stunting pada balita [14]. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap program pencegahan stunting dengan menggunakan metode *Naive Bayes* dan *Support Vector Machine*. Hasil klasifikasi ini diharapkan dapat memberikan wawasan yang berharga bagi pemerintah dalam mengevaluasi efektivitas program pencegahan stunting, serta memperbaiki kebijakan dan strategi komunikasi publik berdasarkan tanggapan masyarakat yang lebih luas dan terukur.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Dalam penelitian ini, gambar 1 menunjukkan metode dilakukan dalam beberapa tahapan yaitu pengumpulan data dari berbagai sumber terkait stunting, preprocessing data untuk membersihkan dan mempersiapkan informasi, labelling untuk memberi label pada entitas dalam dataset,

pembobotan kata untuk menentukan relevansi setiap kata, dan evaluasi model dengan menguji metode *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk menganalisis sentimen terkait stunting.

A. Data Collection

Pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan token otentikasi pada aplikasi X yang diperlukan untuk mengakses *Application Programming Interface* (API). Kata kunci pencarian yang digunakan adalah “stunting” dengan rentang waktu dari September 2023 hingga September 2024, dan hasil yang dicari difokuskan pada bahasa Indonesia. Proses pengambilan data menggunakan *library Tweet Harvester*. Dengan *library* ini, akses yang efisien dan terstruktur terhadap data yang dibutuhkan dapat dilakukan. *Tweet Harvester* berperan penting dalam mengumpulkan *tweet* secara otomatis, sehingga mempermudah proses analisis data yang diperlukan. Hasil pengambilan data tersebut disimpan dalam format CSV dengan nama “stunting.csv”. Dataset yang berhasil dikumpulkan terdiri dari 5907 *tweet*. Setelah proses pengumpulan selesai, file CSV dibaca menggunakan pustaka *pandas* untuk mengelola data. Dari *DataFrame* yang dihasilkan, kolom-kolom penting seperti tanggal pembuatan *tweet*, nama pengguna, dan teks *tweet* diekstraksi. Terakhir, data yang telah dipilih disimpan kembali ke dalam file *stunting.csv* untuk digunakan dalam analisis lebih lanjut.

B. Preprocessing Data

Preprocessing data dalam analisis teks melibatkan beberapa langkah penting. Pertama, *cleaning data* dilakukan untuk menghilangkan karakter-karakter yang tidak relevan, seperti tanda baca, URL, dan symbol-simbol yang tidak diperlukan, sehingga data menjadi lebih bersih dan siap untuk analisis [15]. Selanjutnya, *case folding* dilakukan untuk menyamakan semua huruf menjadi huruf kecil, sehingga konsistensi teks terjaga. [16]. Setelah itu, *tokenizing* memecah teks menjadi kata-kata atau frasa untuk memudahkan analisis. Kemudian, proses *stopword* mengeliminasi kata-kata umum yang tidak memberikan makna signifikan, seperti "dan", "atau", dan "yang". Terakhir, proses *stemming* mengubah kata-kata ke bentuk dasarnya. Dengan langkah-langkah ini, data teks siap untuk dianalisis lebih lanjut.

C. Labelling

Labeling data merupakan proses menambahkan label atau tag pada setiap data dalam sebuah dataset, berdasarkan karakteristik atau atribut yang dimiliki. Dalam analisis sentimen, proses ini digunakan untuk mengklasifikasikan ulasan terkait stunting dengan bantuan hasil evaluasi sentiment [17]. Dataset yang digunakan terdiri dari 5907 *tweet* yang diperoleh dari aplikasi X dan diproses menggunakan *Python* di *Google Colab* untuk memastikan akurasi pengkategorian sentimen.

D. Naïve Bayes

Metode *Naïve Bayes* adalah salah satu algoritma paling sederhana namun kuat dalam klasifikasi berbasis probabilitas, yang menggunakan pendekatan statistik berdasarkan *Teorema Bayes*. *Teorema* ini memberikan dasar matematis untuk menghitung probabilitas bersyarat, atau peluang terjadinya suatu kejadian berdasarkan pengetahuan sebelumnya. *Naïve Bayes* digunakan secara luas dalam masalah klasifikasi, seperti penyaringan spam, analisis sentimen, dan klasifikasi teks, di mana tujuan utamanya adalah mengklasifikasikan suatu data ke dalam kategori tertentu berdasarkan fitur-fitur yang menyertainya [18]. Rumus *Naïve Bayes* berdasarkan teorema Bayes dapat dinyatakan sebagai berikut:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (1)$$

$P(Y|X)$ adalah probabilitas posterior, atau probabilitas bahwa hipotesis X (kelas) benar diberikan data atau bukti Y (fitur). $P(X|Y)$ adalah probabilitas *likelihood*, yaitu probabilitas bahwa data X terjadi jika hipotesis Y benar. $P(Y)$ adalah probabilitas a priori dari hipotesis Y atau seberapa mungkin kelas tersebut sebelum mempertimbangkan bukti baru. $P(X)$ adalah probabilitas dari data X atau probabilitas total dari semua kemungkinan kelas.

Dalam analisis sentimen, *Naïve Bayes* menghitung probabilitas suatu kelas (positif, negatif, atau netral) berdasarkan kata-kata yang muncul dalam teks. Dengan menghitung frekuensi kata-kata dalam masing-masing kelas, model ini dapat menentukan kelas mana yang memiliki kemungkinan tertinggi berdasarkan kata-kata yang ada dalam *tweet*. Keunggulan *Naïve Bayes* terletak pada kesederhanaan dan kecepatan komputasi, sehingga sangat efisien untuk dataset besar, seperti kumpulan *tweet* dalam penelitian ini.

E. Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran mesin yang handal untuk menyelesaikan permasalahan klasifikasi maupun regresi. SVM memiliki beberapa kelebihan dalam analisis stunting, antara lain akurasi yang tinggi, kemampuan mengatasi data non-linier dengan menggunakan kernel trick, serta ketahanan terhadap overfitting, terutama pada dataset kecil. SVM juga efektif dalam menangani data dengan dimensi tinggi, fleksibel dengan berbagai fungsi kernel yang dapat disesuaikan dengan karakteristik data, dan memiliki kemampuan generalisasi yang baik pada data yang belum terlihat. Selain itu, hasil analisis SVM, seperti support vectors, memberikan wawasan tentang faktor-faktor kunci yang mempengaruhi stunting, menjadikannya alat yang efektif untuk analisis dan klasifikasi dalam konteks kesehatan masyarakat [19]. Algoritma ini bekerja dengan mencari batas keputusan (*hyperplane*) yang paling baik untuk

memisahkan data menjadi kelompok-kelompok yang berbeda berdasarkan karakteristik fitur. Pemilihan *hyperplane* didasarkan pada prinsip memaksimalkan margin, yaitu jarak terdekat antara *hyperplane* dan data dari masing-masing kelas. Berikut rumusnya :

$$f(x) = \text{sign}(w^T x + b) \quad (2)$$

w merupakan Vektor bobot yang menentukan arah dan kemiringan *hyperplane*. Bobot ini diperoleh selama proses pelatihan model. x adalah Vektor fitur dari data baru yang ingin diklasifikasikan. b merupakan Bias yang menyesuaikan posisi *hyperplane* dalam ruang fitur. Bias ini juga dihasilkan selama pelatihan. $w^T x + b$ adalah fungsi linear yang menghitung jarak data x dari *hyperplane*. Hasilnya bisa positif, negatif, atau nol. $\text{sign}()$ fungsi ini mengembalikan tanda dari nilai yang dihitung. Jika hasilnya lebih besar dari nol, fungsi ini mengembalikan +1 (menunjukkan bahwa data tersebut termasuk dalam kelas positif). Jika hasilnya kurang dari nol, fungsi ini mengembalikan -1 (menunjukkan bahwa data tersebut termasuk dalam kelas negatif).

F. Evaluation Model

Evaluasi model dilakukan untuk mengukur akurasi dan kinerja dari kedua metode (*Naïve Bayes* dan SVM) dalam mengklasifikasikan sentiment [20]. Proses ini mencakup pengujian menggunakan metrik seperti akurasi, presisi, recall, dan F1-Score untuk menilai efektivitas masing-masing model.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1 Pembagian Data

Dalam tahap pengumpulan data, penelitian ini memperoleh total 5907 tweet yang berkaitan dengan topik yang dianalisis. Setelah data terkumpul, langkah selanjutnya adalah membagi dataset tersebut dengan rasio 80:20 untuk keperluan pelatihan dan pengujian model. Pemilihan rasio 80:20 dalam pembagian data dilakukan untuk memastikan bahwa model memiliki cukup data untuk dilatih, yaitu sebanyak 80% dari total dataset, sementara 20% sisanya digunakan sebagai data pengujian. Rasio ini memberikan keseimbangan yang baik antara pelatihan dan pengujian, sehingga model dapat mempelajari pola secara efektif dari data training yang lebih besar, tetapi tetap diuji pada data yang belum pernah dilihat sebelumnya untuk mengukur kemampuannya dalam menggeneralisasi. Sebanyak 80% dari total data, yaitu 4725 tweet, digunakan

sebagai data training untuk melatih model, sedangkan 20% sisanya, sebanyak 1182 tweet, dialokasikan sebagai data testing untuk mengukur performa model. Pembagian data ini dilakukan agar model dapat diukur kemampuannya dalam menggeneralisasi pada data yang belum pernah dilihat. Dari keseluruhan tweet, terdapat 5570 tweet yang diklasifikasikan sebagai netral, 226 tweet positif, dan 111 tweet negatif, mencerminkan distribusi sentimen yang akan digunakan dalam proses klasifikasi lebih lanjut.

3.2 Preprocessing Data

a. Cleaning Data

Cleaning data adalah proses membersihkan data dari kesalahan-kesalahan umum seperti data yang hilang, duplikat, atau format yang tidak konsisten. Tujuannya adalah untuk mendapatkan data yang bersih dan siap untuk dianalisis [21]. Hasil pembersihan data terlihat pada tabel 1.

Tabel 1. *Cleaning Data*

<i>Tweet</i>	<i>Cleaning Data</i>
Ayo Dukung! Pilih produk yang difortifikasi dan edukasi orang di sekitar kamu tentang pentingnya pangan bergizi. Bersama kita bisa turunkan angka stunting di Indonesia! Safer Food Better Health #Stunting #PanganFortifikasi #IndonesiaSehat #BPOM #Fortifikasi https://t.co/kraVb3JyTC	Ayo Dukung Pilih produk yang difortifikasi dan edukasi orang di sekitar kamu tentang pentingnya pangan bergizi Bersama kita bisa turunkan angka stunting di Indonesia Safer Food Better Health Stunting PanganFortifikasi IndonesiaSehat BPOM Fortifikasi
@mas_veel Bank dunia harusnya tau diri.. MAkan siang gratis di Indonesia itu untuk mengatasi stunting otak para pemimpin di Indonesia...rakyatnya dah sehat semua...sehat mental dan jiwa...kalau di Indonesia yg kena stunting mental dan jiwa itu para pemimpinnya	Bank dunia harusnya tau diri MAkan siang gratis di Indonesia itu untuk mengatasi stunting otak para pemimpin di Indonesia rakyatnya dah sehat semua sehat mental dan jiwa kalau di Indonesia yg kena stunting mental dan jiwa itu para pemimpinnya

b. Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. Langkah awal yang umum dalam pra-pemrosesan teks (*text preprocessing*) sebelum melakukan analisis teks lebih lanjut, seperti stemming atau analisis sentimen [22]. Hasil dari *case folding* terlihat pada tabel 2.

Tabel 2. *Case Folding*

<i>Tweet</i>	<i>Case Folding</i>
Ayo Dukung Pilih produk yang difortifikasi dan edukasi orang di sekitar kamu tentang pentingnya pangan bergizi Bersama kita bisa turunkan angka stunting di Indonesia Safer Food Better Health Stunting PanganFortifikasi IndonesiaSehat BPOM Fortifikasi	ayo dukung pilih produk yang difortifikasi dan edukasi orang di sekitar kamu tentang pentingnya pangan bergizi bersama kita bisa turunkan angka stunting di indonesia safer food better health stunting panganfortifikasi indonesiasehat bpom fortifikasi

Bank dunia harusnya tau diri
MAkan siang gratis di
Indonesia itu untuk
mengatasi stunting otak para
pemimpin di Indonesia
rakyatnya dah sehat semua
sehat mental dan jiwa kalau
di Indonesia yg kena stunting
mental dan jiwa itu para
pemimpinnya

bank dunia harusnya tau diri
makan siang gratis di
indonesia itu untuk
mengatasi stunting otak para
pemimpin di indonesia
rakyatnya dah sehat semua
sehat mental dan jiwa kalau
di indonesia yg kena stunting
mental dan jiwa itu para
pemimpinnya

c. *Tokenizing*

Tokenisasi adalah proses memecah teks menjadi unit-unit individual yang disebut token. Token ini dapat berupa kata, kalimat, frasa, atau bahkan karakter, tergantung pada aplikasi tertentu. Tokenisasi adalah langkah mendasar dalam banyak tugas pemrosesan bahasa alami (NLP), seperti klasifikasi teks, pencarian informasi, dan terjemahan mesin [23]. Hasil dari *tokenizing* terlihat pada tabel 3.

Tabel 3. *Tokenizing*

<i>Tweet</i>	<i>Tokenizing</i>
ayo dukung pilih produk yang difortifikasi dan edukasi orang di sekitar kamu tentang pentingnya pangan bergizi bersama kita bisa turunkan angka stunting di indonesia safer food better health stunting panganfortifikasi indonesia sehat bpom fortifikasi	['', 'ayo', 'dukung', 'pilih', 'produk', 'yang', 'difortifikasi', 'dan', 'edukasi', 'orang', 'di', 'sekitar', 'kamu', 'tentang', 'pentingnya', 'pangan', 'bergizi', 'bersama', 'kita', 'bisa', 'turunkan', 'angka', 'stunting', 'di', 'indonesia', 'safer', 'food', 'better', 'health', 'stunting', 'panganfortifikasi', 'indonesiasehat', 'bpom', 'fortifikasi', '']
bank dunia tau makan siang gratis indonesia mengatasi stunting otak pemimpin indonesia rakyatnya dah sehat sehat mental jiwa indonesia yg kena stunting mental jiwa pemimpinnya	['', 'bank', 'dunia', 'harusnya', 'tau', 'diri', 'makan', 'siang', 'gratis', 'di', 'indonesia', 'itu', 'untuk', 'mengatasi', 'stunting', 'otak', 'para', 'pemimpin', 'di', 'indonesia', 'rakyatnya', 'dah', 'sehat', 'sehat', 'mental', 'dan', 'jiwa', 'kalau', 'di', 'indonesia', 'yg', 'kena', 'stunting', 'mental', 'dan', 'jiwa', 'itu', 'para', 'pemimpinnya', '']

d. *Stopword*

Stopword removal adalah proses eliminasi kata-kata umum yang tidak berkontribusi pada makna semantik teks, sehingga analisis dapat lebih terfokus pada kata-kata kunci. Hasil dari *stopword* terlihat pada tabel 4.

Tabel 4. *StopWord*

<i>Tweet</i>	<i>StopWord</i>
['', 'ayo', 'dukung', 'pilih', 'produk', 'yang', 'difortifikasi', 'dan', 'edukasi', 'orang', 'di', 'sekitar', 'kamu', 'tentang', 'pentingnya', 'pangan', 'bergizi', 'bersama', 'kita', 'bisa', 'turunkan', 'angka', 'stunting', 'di', 'indonesia', 'safer', 'food', 'better', 'health', 'stunting', 'panganfortifikasi', 'indonesiasehat', 'bpom', 'fortifikasi', '']	ayo dukung pilih produk difortifikasi edukasi orang pangan bergizi turunkan angka stunting indonesia safer food better health stunting panganfortifikasi indonesiasehat bpom fortifikasi
['', 'bank', 'dunia', 'harusnya', 'tau', 'diri', 'makan', 'siang', 'gratis', 'di', 'indonesia', 'itu', 'untuk', 'mengatasi', 'stunting', 'otak', 'para', 'pemimpin', 'di', 'indonesia', 'rakyatnya', 'dah', 'sehat', 'sehat', 'mental', 'dan', 'jiwa', 'kalau', 'di', 'indonesia', 'yg', 'kena', 'stunting', 'mental', 'dan', 'jiwa', 'itu', 'para', 'pemimpinnya', '']	bank dunia tau makan siang gratis indonesia mengatasi stunting otak pemimpin indonesia rakyatnya dah sehat

'sehat', 'semua', 'sehat', 'mental', sehat mental jiwa
'dan', 'jiwa', 'kalau', 'di', 'indonesia', indonesia yg kena
'yg', 'kena', 'stunting', 'mental', 'dan', stunting mental jiwa
'jiwa', 'itu', 'para', 'pemimpinnya', ''] pemimpinnya

e. *Stemming*

Stemming merupakan salah satu teknik pra-pemrosesan teks yang penting dalam pemrosesan bahasa alami. Teknik ini digunakan untuk menyederhanakan teks, sehingga dapat meningkatkan efisiensi dan akurasi algoritma pembelajaran mesin. Hasil dari *stemming* terlihat pada tabel 5.

Tabel 5. *Stemming*

<i>Tweet</i>	<i>Stemming</i>
['', 'ayo', 'dukung', 'pilih', 'produk', 'yang', 'difortifikasi', 'dan', 'edukasi', 'orang', 'di', 'sekitar', 'kamu', 'tentang', 'pentingnya', 'pangan', 'bergizi', 'bersama', 'kita', 'bisa', 'turunkan', 'angka', 'stunting', 'di', 'indonesia', 'safer', 'food', 'indonesia', 'safer', 'food', 'better', 'health', 'stunting', 'panganfortifikasi', 'indonesiasehat', 'bpom', 'fortifikasi', '']	['', 'ayo', 'dukung', 'pilih', 'produk', 'yang', 'fortifikasi', 'dan', 'edukasi', 'orang', 'di', 'sekitar', 'kamu', 'tentang', 'penting', 'pangan', 'gizi', 'sama', 'kita', 'bisa', 'turun', 'angka', 'stunting', 'di', 'indonesia', 'safer', 'food', 'better', 'health', 'stunting', 'panganfortifikasi', 'indonesiasehat', 'bpom', 'fortifikasi', '']
['', 'bank', 'dunia', 'harusnya', 'tau', 'diri', 'makan', 'siang', 'gratis', 'di', 'indonesia', 'itu', 'untuk', 'mengatasi', 'stunting', 'otak', 'para', 'pemimpin', 'di', 'indonesia', 'rakyatnya', 'dah', 'sehat', 'sehat', 'mental', 'dan', 'jiwa', 'semua', 'sehat', 'mental', 'kalau', 'di', 'indonesia', 'yg', 'kena', 'stunting', 'mental', 'dan', 'jiwa', 'itu', 'para', 'pimpin', '']	['', 'bank', 'dunia', 'harus', 'tau', 'diri', 'makan', 'siang', 'gratis', 'di', 'indonesia', 'itu', 'untuk', 'atas', 'stunting', 'otak', 'para', 'pimpin', 'di', 'indonesia', 'rakyat', 'dah', 'sehat', 'semua', 'sehat', 'mental', 'dan', 'jiwa', 'kalau', 'di', 'indonesia', 'yg', 'kena', 'stunting', 'mental', 'dan', 'jiwa', 'itu', 'para', 'pimpin', '']

Pada gambar 6 adalah kalimat yang lebih jelas untuk menjelaskan situasi data training dan testing sebelum pembobotan kata atau penerapan *TF-IDF*:

Tabel 6. Hasil Data *Training* dan *Data Testing*

<i>Hasil Data Training</i>	<i>Hasil Data Testing</i>
2091 statistik indonesia tingkat stunting gizi bur...	5369 keren banget yaaa langsung menargetkan permasa...
1234 selamat keluarga nasional keluarga bebas stunt...	978 segitu persentase stunting indo lo nyalahin ko...
169 solusi stunting pra kehamilan kehamilan menen...	3008 lucinta luna jiafei oplasnya dewa bgt tpi kalo...
5220 makan siang susu gratis wacana rencana aksi pr...	5338 program makan siang susu gratis mencegah stunt...
5300 bener untik cegah stunting bagus nutrisinya ma...	

3.3 *Classification Model*

Model klasifikasi yang digunakan dalam penelitian ini dievaluasi melalui perbandingan antara hasil prediksi model dan data aktual. Sebelum evaluasi dilakukan, dataset dibagi menjadi

dua bagian: data training, yang digunakan untuk melatih model, dan data testing, yang berfungsi untuk menguji kinerja model. Pembagian data dilakukan dengan rasio 80:20, sehingga menciptakan keseimbangan yang optimal antara pelatihan dan pengujian, dan memastikan bahwa model dapat diukur secara akurat dalam memprediksi hasil pada data yang belum pernah dilihat sebelumnya.

Tabel 7. Hasil Perbandingan *Data Training* dan *Data Testing*

Modelling	Data	Data
	Training	Testing
Support Vector Machine	96,91%	95,43%
Naïve Bayes	95,34%	84,52%

Pada tabel 7 hasil pengujian model klasifikasi, algoritma SVM menunjukkan performa yang baik dibandingkan dengan *Naïve Bayes*. SVM mencatatkan akurasi sebesar 96,91% pada data training dan 95,43% pada data testing, yang menunjukkan kemampuannya dalam generalisasi terhadap data yang belum terlihat. Sementara itu, *Naïve Bayes* memperoleh akurasi 95,34% pada data training, tetapi mengalami penurunan signifikan pada data testing dengan akurasi hanya 84,52%. Hal ini mengindikasikan bahwa meskipun *Naïve Bayes* dapat memberikan hasil yang baik pada data yang digunakan untuk pelatihan, model ini kurang efektif dalam memprediksi data baru, sehingga memerlukan perhatian lebih dalam hal pengembangan dan optimisasi model untuk meningkatkan kinerjanya pada data testing.

Tabel 8. Hasil *Confusion Matrix Data Testing Modelling*

Modelling	Accuracy	Precision	Recall
Support Vector Machine	95,43%	98,33%	44,00%
Naïve Bayes	84,52%	38,00%	41,00%

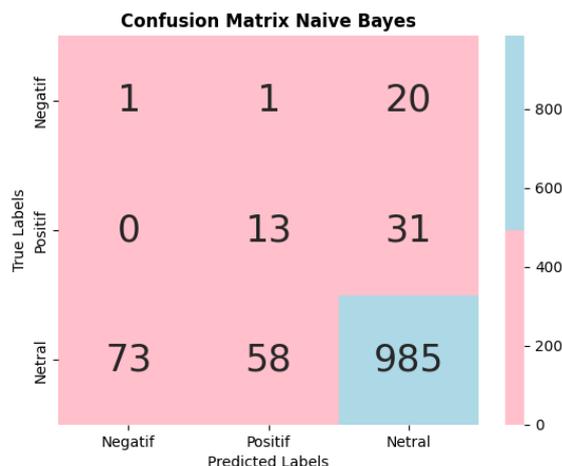
Tabel 8 menunjukkan hasil *Confusion Matrix* untuk data testing dari kedua model yang digunakan. Model SVM mencapai akurasi 95,43% dengan precision 98,33% dan recall 44,00%, menunjukkan kinerja yang lebih baik dibandingkan dengan model *Naïve Bayes* yang memiliki akurasi 84,52%, precision 38,00%, dan recall 41,00%. Rendahnya nilai precision pada model *Naïve Bayes* mengindikasikan banyaknya prediksi positif yang salah, yang dapat mempengaruhi keandalan model dalam klasifikasi. Sementara itu, meskipun recall pada SVM lebih tinggi, angka yang lebih rendah menunjukkan bahwa model tersebut masih perlu meningkatkan kemampuan dalam mengidentifikasi semua kasus positif dengan lebih akurat.

3.4 Confusion Matrix

a. Confusion Matrix Naïve Bayes

Pada gambar 4 *Confusion Matrix* digunakan untuk mengevaluasi kinerja model klasifikasi

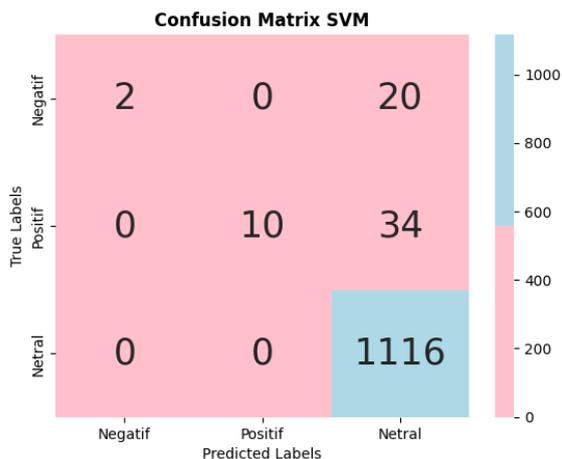
dengan membandingkan prediksi model terhadap data actual [24]. model berhasil memprediksi kelas negatif dengan benar sebanyak 1 kali (*True Negative*), namun salah memprediksi negatif sebagai positif sebanyak 1 kali (*False Positive*), serta salah memprediksi negatif sebagai netral sebanyak 20 kali (*False Negative*). Untuk kelas positif, model melakukan 13 prediksi yang benar (*True Positive*), tetapi salah memprediksi positif sebagai netral sebanyak 31 kali (*False Positive*). Sementara itu, untuk kelas netral, model berhasil melakukan 985 prediksi yang benar (*True Neutral*), tetapi salah memprediksi netral menjadi negatif sebanyak 73 kali dan prediksi netral menjadi positif sebanyak 58 kali. Hasil ini menunjukkan bahwa model perlu perbaikan dalam memprediksi kelas negatif dan positif, meskipun performa pada kelas netral cukup baik.



Gambar 4. *Confusion Matrix Naive Bayes*

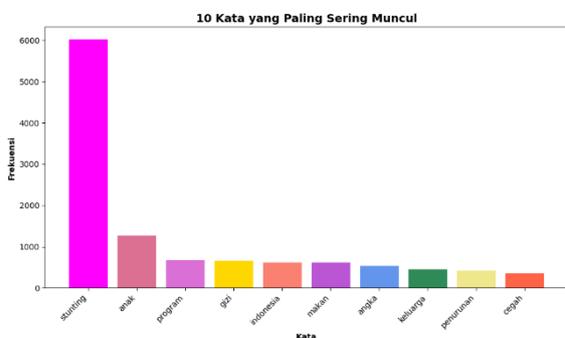
b. Confusion Matrix Support Vector Machine

Berdasarkan data pada gambar 5 dari *Confusion Matrix*, model berhasil memprediksi kelas negatif dengan benar sebanyak 2 kali (*True Negative*), tetapi salah memprediksi negatif sebagai netral sebanyak 20 kali (*False Negative*). Untuk kelas positif, model melakukan 10 prediksi yang benar (*True Positive*), namun salah memprediksi positif sebagai netral sebanyak 34 kali (*False Positive*). Sementara itu, untuk kelas netral, model berhasil melakukan 1116 prediksi yang benar (*True Neutral*), tanpa kesalahan prediksi. Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam mengidentifikasi kelas netral, tetapi perlu perbaikan dalam memprediksi kelas negatif dan positif untuk meningkatkan akurasi keseluruhan model.



Gambar 5. Confusion Matrix SVM

3.5 Visualisasi Data dan WordCloud



Gambar 6. Kata yang paling sering muncul

Berdasarkan gambar 6 menunjukkan hasil analisis frekuensi kata, terlihat bahwa kata “stunting” muncul paling sering, yaitu sebanyak 6020 kali, menandakan bahwa isu ini menjadi topik utama dalam diskusi atau dokumen yang dianalisis. Kata “anak”, yang muncul sebanyak 1269 kali, menunjukkan fokus utama pada kelompok anak-anak yang menjadi target dari berbagai upaya terkait stunting. Selain itu, kata “program” 679 kali dan “gizi” 665 kali mengindikasikan adanya banyak pembahasan mengenai intervensi dan pentingnya pemenuhan nutrisi dalam mengatasi stunting di Indonesia. Kata “Indonesia” muncul sebanyak 620 kali, mempertegas bahwa stunting menjadi tantangan kesehatan nasional. Frekuensi tinggi kata “makan” 617 kali, menunjukkan bahwa pola makan menjadi aspek penting dalam diskusi ini. Sementara itu, kata “angka” yang muncul 536 kali menandakan adanya perhatian pada data dan statistik untuk memantau perkembangan atau dampak dari stunting. Kata-kata seperti “keluarga”, “penurunan”, dan “cegah” dengan frekuensi masing-masing 449, 429, dan 353 kali menunjukkan pola komunikasi informal yang digunakan dalam diskusi terkait stunting.



Gambar 7. WordCloud Stunting

Berdasarkan hasil dari gambar wordcloud 7 yang menampilkan kata terbesar “stunting”, dapat diasumsikan bahwa topik utama yang sering dibahas adalah terkait dengan program penurunan angka stunting di masyarakat. Selain itu, keberadaan program untuk menurunkan angka stunting menjadi salah satu fokus utama dalam upaya mencegah stunting secara lebih luas.

4. DISKUSI

Penelitian ini menunjukkan analisis sentimen terkait program pencegahan stunting di Indonesia menunjukkan bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan dengan *Naive Bayes*, dengan akurasi SVM mencapai 95,43% dan *Naive Bayes* sebesar 84,52%. Hal ini mengindikasikan bahwa SVM lebih efektif dalam mengklasifikasikan sentimen publik mengenai topik tersebut.

Hasil ini sejalan dengan penelitian sebelumnya bahwa SVM memberikan performa akurasi tinggi dalam klasifikasi data kesehatan, yaitu sebesar 82% pada kasus stunting balita, dengan presisi 80% dan recall 86% [8]. Sebaliknya, penelitian Supriyadi (2022) dalam mendeteksi *Autism Spectrum Disorder* (ASD) justru menemukan bahwa *Naive Bayes* lebih unggul dengan akurasi 96,45%, sementara SVM hanya mencapai 81,56%. Perbedaan hasil ini menunjukkan bahwa efektivitas algoritma dalam klasifikasi dapat bergantung pada jenis data dan konteks kasus yang dianalisis [7].

Hal ini mendukung temuan bahwa SVM memiliki keunggulan dalam menangani data yang lebih kompleks dan menghasilkan klasifikasi yang lebih akurat pada beberapa jenis data. Dengan demikian, penelitian ini menekankan bahwa baik SVM maupun *Naive Bayes* memiliki keunggulan masing-masing, tergantung pada karakteristik data yang dianalisis. Untuk analisis sentimen program pemerintah seperti stunting, SVM tampaknya lebih efektif, namun *Naive Bayes* tetap relevan pada kasus tertentu, seperti deteksi ASD dan klasifikasi stunting berbasis probabilistik. Penggunaan data media sosial dan machine learning juga semakin penting dalam memahami persepsi masyarakat dan memandu pengambilan keputusan kebijakan.

5. KESIMPULAN

Berdasarkan penelitian yang dilakukan, program pencegahan stunting di Indonesia mendapat respons beragam dari masyarakat. Melalui analisis sentimen dengan metode *Naïve Bayes* dan SVM, ditemukan bahwa algoritma SVM menunjukkan kinerja yang lebih baik dalam pengujian dengan akurasi 95,43%, dibandingkan dengan *Naïve Bayes* yang hanya mencapai 84,52% pada data testing, meskipun *Naïve Bayes* unggul dalam pelatihan dengan akurasi 95,34%. Hasil ini menunjukkan bahwa algoritma SVM lebih efektif dalam beradaptasi dengan data yang kompleks, berkat kemampuannya untuk menemukan batas keputusan optimal melalui pemetaan fitur dan fungsi kernel. Sebaliknya, meskipun *Naïve Bayes* cepat dan mudah diinterpretasikan, algoritma ini rentan terhadap *noise* dan bergantung pada asumsi independensi antara fitur, yang membatasi kemampuannya dalam menangani hubungan yang lebih rumit dalam data. Selain itu, analisis dataset mengidentifikasi sepuluh kata yang paling sering muncul, dengan 'stunting' sebagai kata yang paling dominan 6020 kali, diikuti oleh 'anak' 1269 kali. Hasil *wordcloud* dan frekuensi kemunculan kata-kata seperti stunting dan anak mencerminkan perhatian terhadap isu stunting di Indonesia. Oleh karena itu, analisis sentimen dapat berfungsi sebagai alat yang berguna untuk memahami persepsi masyarakat serta menilai efektivitas program pemerintah. Penelitian ini juga menekankan pentingnya memanfaatkan data dari media sosial untuk menggali opini publik, sehingga langkah-langkah perbaikan yang tepat dapat diambil guna meningkatkan keberhasilan program pencegahan stunting.

SARAN

Untuk penelitian selanjutnya, disarankan agar dilakukan pelabelan manual melalui anotasi data. Dengan cara ini, diharapkan akurasi model dapat meningkat dan dataset menjadi lebih lengkap, sehingga analisis yang dihasilkan dapat lebih tepat dan bermanfaat.

DAFTAR PUSTAKA

- [1] I. A. Parmitha, R. Arifinia, G. K. Pangestu, N. A. Rahayu, and A. Rosidi, "Gambaran Kejadian Stunting Berdasarkan Karakteristik Ibu Pada Balita Usia 24-59 Bulan," *J. Inov. Ris. Ilmu Kesehat.*, vol. 15, no. 1, pp. 37–48, 2024.
- [2] Y. Haskas, "Gambaran Stunting Di Indonesia: Literatur Review," *J. Ilm. Kesehat. Diagnosis*, vol. 15, no. 2, pp. 2302–2531, 2020.
- [3] F. N. Zaman, M. A. Fadhilah, M. A. Ulinuha, and K. Umam, "Menganalisis Respons Netizen Twitter Terhadap Program Makan Siang Gratis Menerapkan Nlp Metode Naïve Bayes," *J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 14, no. 3, pp. 150–233, 2024, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [4] R. R. S. Putri Kumala Sari, "Komparasi Algoritma Support Vector Machine dan Random Forest Untuk Analisis Sentimen Metaverse," *J. Mnemon.*, vol. 7, no. 1, pp. 31–39, 2024.
- [5] Eskiyaturrofikoh and R. R. Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm)," vol. 8, no. 3, pp. 110–118, 2019, doi: <https://doi.org/10.29100/jipi.v9i3.5392>.
- [6] Ramadhani, Ramadhanu, and Taufik Hidayat, "Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree," *J. SIMETRIS*, vol. 15, no. 1, pp. 57–68, 2024, [Online]. Available: <https://jurnal.umk.ac.id/index.php/simet/article/view/10679>
- [7] R. Supriyadi, N. Maulidah, A. Fauzi, H. Nalatissifa, and S. Diantika, "Penerapan Algoritma Naïve Bayes dan Support Vector Machine dalam Memprediksi Autisme," *J. Swabumi*, vol. 10, no. 1, p. 2022, 2022, doi: <https://doi.org/10.31294/swabumi.v10i1.12294>.
- [8] F. O. Dayera, Musa Bundaris Palungan, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, doi: [10.33379/gtech.v8i3.4811](https://doi.org/10.33379/gtech.v8i3.4811).
- [9] Y. Mulyanto, F. Idifitriani, A. Wati, U. T. Sumbawa, D. Mining, and K. P. Tano, "Klasifikasi Data Mining Untuk Penentuan Stunting Pada Balita Menggunakan Metode Naïve Bayes," *J. Mnemon.*, vol. 7, no. 2, pp. 119–125, 2024.
- [10] A. Setiawan and R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: [10.29408/edumatic.v8i1.25667](https://doi.org/10.29408/edumatic.v8i1.25667).
- [11] Y. Yusra and M. Fikry, "Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 50, 2018, doi: [10.24014/coreit.v4i2.5205](https://doi.org/10.24014/coreit.v4i2.5205).
- [12] R. Mursyid, A. D. Indriyanti, J. Teknik, I. Fakultas, T. Universitas, and N. Surabaya, "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna

- Pada Platform Media Sosial (Studi Kasus : Twitter),” vol. 06, pp. 371–383, 2024.
- [13] W. Ningsih, B. Alfianda, Rahmaddeni, and D. Wulandari, “Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” vol. 4, no. 2, pp. 556–562, 2024.
- [14] Harliana and D. Anggraini, “Penerapan Algoritma Naive Bayes Pada Klasifikasi Status Gizi Balita di Posyandu Desa Kalitengah,” *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 21, no. 2, pp. 38–45, 2023, doi: 10.61805/fahma.v21i2.16.
- [15] A. Sudrajat, N. Mulyani, and N. Marpaung, “Sistem Pendukung Keputusan Penentuan Kelayakan Penangguhan Kredit Nasabah menggunakan Naive Bayes,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 2, pp. 205–214, 2022, doi: 10.29408/edumatic.v6i2.6298.
- [16] R. Fatmasari, R. Septiani, T. Pinem, D. Fabiyanto, and W. Gata, “Implementasi Algoritma BERT Pada Komentar Layanan Akademik dan Non Akademik Universitas Terbuka di Media Sosial,” *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 5, p. 96, Jan. 2024, doi: 10.30872/jsakti.v5i2.13915.
- [17] A. S. Pamungkas and N. Cahyono, “Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 1–10, 2024, doi: 10.29408/edumatic.v8i1.24114.
- [18] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, “Comparison of Support Vector Machine and Naive Bayes on Twitter Data Sentiment Analysis,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.
- [19] D. Atmajaya, A. Febrianti, and H. Darwis, “Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter.” *Indones. J. Comput. Sci.*, vol. 12, Aug. 2023, doi: 10.33022/ijcs.v12i4.3341.
- [20] S. Lestari and S. Berliani, “Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine,” vol. 5, no. 3, pp. 951–960, 2024, doi: <https://doi.org/10.55338/saintek.v5i3.2746>.
- [21] M. U. Albab, Y. Karuniawati P, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” *J. Transform.*, vol. 20, no. 2, pp. 1–10, 2023, [Online]. Available: <https://journals.usm.ac.id/index.php/transformatika/>page1
- [22] Sri Diantika, Windu Gata, and Hiya Nalattissifa, “Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik,” *Elkom J. Elektron. dan Komput.*, vol. 14, no. 1, pp. 10–15, 2021, doi: 10.51903/elkom.v14i1.319.
- [23] Rina Noviana and Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [24] Z. Amri, K. Kusriani, and K. Kusnawi, “Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes, Decision Tree, ANN, KNN, dan SVM,” *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 187–196, 2023, doi: 10.29408/edumatic.v7i2.18620.